

DOCUMENT RESUME

ED 471 988

TM 034 700

AUTHOR Thomas, John P.  
TITLE An Examination of the Relationship between Learning Variables and Academic Achievement Outcomes.  
PUB DATE 2002-12-00  
NOTE 33p.  
PUB TYPE Reports - Research (143)  
EDRS PRICE EDRS Price MF01/PC02 Plus Postage.  
DESCRIPTORS \*Academic Achievement; \*Junior High School Students; Junior High Schools; Mathematics; Multivariate Analysis; National Surveys; \*Productivity; Reading; Sciences; \*Student Characteristics  
IDENTIFIERS \*National Education Longitudinal Study 1988

ABSTRACT

Using eighth-grade transcript and questionnaire data from the National Education Longitudinal Study of 1988, the current study used multivariate procedures to examine the nature of the relationship between educational productivity factors and academic achievement outcomes in reading, mathematics, and science in a sample of approximately 450 eighth graders. Using canonical correlation, discriminant analysis, factor analysis, and multivariate analysis of variance procedures, the results of the current study showed that approximately 40% of the variance in achievement outcomes is attributable to productivity factors. Mathematics achievement was influenced most by advanced course work in mathematics, reading was most influenced by grades in English courses, and both reading and science are strongly affected by parent's education level. In addition, grades, family background, and time spent on homework had an impact on achievement outcomes in all three areas. (Author/SLD)

# An Examination of the Relationship Between Learning Variables and Academic Achievement Outcomes

**Dr. John P. Thomas**

College of Lake County  
Grayslake, Illinois

U.S. DEPARTMENT OF EDUCATION  
Office of Educational Research and Improvement  
EDUCATIONAL RESOURCES INFORMATION  
CENTER (ERIC)

- ☒ This document has been reproduced as received from the person or organization originating it.
- ☐ Minor changes have been made to improve reproduction quality.
- Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE AND  
DISSEMINATE THIS MATERIAL HAS  
BEEN GRANTED BY

**J. Thomas**

TO THE EDUCATIONAL RESOURCES  
INFORMATION CENTER (ERIC)

1

## **Abstract**

Using eighth-grade transcript and questionnaire data from the National Education Longitudinal Study of 1988, the current study used multivariate procedures to examine the nature of the relationship between educational productivity factors and academic achievement outcomes in reading, math, and science. Using canonical correlation, discriminant analysis, factor analysis, and multivariate analysis of variance procedures, the results of the current study showed that approximately 40% of the variance in achievement outcomes is attributable to productivity factors. Math achievement was influenced most by advanced course work in mathematics, reading was most influenced by grades in English courses, and both reading and science are strongly effected by parent's education level. Additionally, grades, family background, and time spent on homework had an impact on achievement outcomes in all three areas.

## **Introduction**

Previous studies have examined the effect of educational productivity factors on achievement outcomes. Most of these analyses have however focused on a single subject area for which the productivity factors are associated. Very few studies have attempted to examine the relationships between the factors and multiple dependent measures.

To consider the association between multiple dependent variables and a set of independent variables, multivariate statistical techniques such as canonical correlation analysis, discriminant analysis, and multivariate analysis of variance (MANOVA) can be applied. In this particular study, an attempt was made to determine the “best” models to evaluate the linear relationships between educational productivity factors and proficiency in three academic subject areas: mathematics, reading, and science. The proficiency scores are not explicitly continuous, however previous studies have suggested that they perform similarly to raw or standardized scores or more traditional continuous measures of achievement (U.S. Department of Education, 1994).

### **Purpose of the Study**

The purpose of the current study is to determine the nature of the relationship between educational productivity factors and achievement in three academic subject areas: reading, math, and science. The first part of the study uses canonical correlation analysis to determine maximally correlated linear combinations of productivity factors and the proficiency levels in the three subject areas, respectively. The second part of the analysis uses discriminant analysis to determine the most appropriate model consisting of educational productivity factors to distinguish between proficiency and non-proficiency for the 3 subject areas. The third part of the analysis attempts to reduce the dimension of

the model or set of productivity factors in order to gain a better understanding of the interrelationships among the variables via principle component analysis and principle factor analysis. The latent factors determined from this part of the analysis were utilized in the final part of the analysis. In particular, based on the results of the factor analysis procedures, certain subsets of variables were combined to form scales for underlying latent variables. The fourth part of the analysis uses MANOVA to determine whether there are differences in mean achievement in reading, mathematics, and science relative to different levels of the “new” scales as well as between gender and race.

### **Study**

#### **A. Participants**

All of the participants in the present study were eighth grade students who took part in the base year wave of data collection for the National Education Longitudinal Study of 1988 (NELS: 88). The base year of the NELS: 88 “represents the first stage of a major longitudinal effort designed to provide trend data about critical transitions experienced by students as they leave elementary school and progress through high school and into college or their careers” (U.S. Department of Education, 1989, p. 19).

#### **B. Data Collection**

For the base-year component of the NELS: 88, a two-stage stratified probability sample was used to select a nationally representative sample of schools and students. For the first stage, schools constituted the primary sampling unit. A pool of 1,032 schools was selected through stratified sampling with probability of selection proportional to eighth-grade size and with over sampling of private schools. A pool of 1,032 replacement schools was selected by the same method. Of the 1,032 initial selections, 30

were considered ineligible. Of the 1,002 eligible schools, 698 participated. An additional 359 schools (supplied by alternative selections from the replacement pool) also participated, for a total sample of 1,057 cooperating schools, of which 1,052 schools (815 public schools and 237 private schools) contributed usable student data. In addition, to the selection process described above, over-sampling of schools with very large percentages of African-American or Hispanic students or both was conducted based on information provided to the National Opinion Research Center (NORC), a subcontractor for the NELS: 88 base-year study, and by the Office of Civil Rights (ORC) and other sources.

The second stage of the base-year selection process produced a random selection of 26,435 students among the sampled schools, resulting in participation by 24,323 eighth-grade students (an average of 23 students per school). For the current study, the sample was reduced to approximately 10,000 after eliminating those who had incomplete transcript information or either non-response or multiple responses for items used in the current research.

Using random sampling procedures, a sub-sample of approximately 450 students was selected for the present study. The relative frequencies for the item-responses among the sub-sample were very similar to the frequencies obtained from the full sample.

The data for the present study was acquired from the Public Use version of the CD-ROM Electronic Codebook System of the National Education Research Longitudinal Study: 1988-1994 (1996). The data files on this CD-ROM included results from the base-year through third follow-up of the NELS: 88.

### C. Instrument

The base-year study of the NELS: 88 included a self-administered questionnaire for gathering information about background variables and a range of other topics including school work, aspirations, and social relationships (Peng & others, 1995). The students also completed a series of cognitive tests developed by the Educational Testing Service (ETS). The cognitive test battery included a multiple choice mathematics test, which consisted of quantitative comparisons and other questions assessing mathematical knowledge. The student questionnaire asked students about such topics as academic achievement; student perceptions of their curriculum and school, family structure and environment; social relations; and aspirations, attitudes, and values, particularly as they relate to high school and occupational or post-secondary educational plans (U.S. Department of Education, 1994).

### D. Procedure

The present study used multivariate procedures to ascertain the relationships between several background variables (i.e., educational productivity factors) and achievement in reading, math, and science. Canonical correlation analysis and discriminant analysis were used to determine the nature of the relationship between the productivity factors and achievement in the three subject areas. For these procedures, proficiency levels served as a measure of achievement. Factor analysis was applied to the set of scores for the productivity factors to examine whether it was possible to reduce the dimension of this set of variables or create latent constructs for subsequent analyses. Multivariate analysis of variance (MANOVA) was used in the final phase to determine if there were differences in mean achievement in reading, math, and science, respectively

across levels of the productivity factors as well as between gender and race. In the final phase, the students' standardized scores on the cognitive tests in reading, math, and science represented "achievement" level. All of the statistical analyses used in the current study, including the recoding and construction of composite measures, were done using SAS software.

#### E. Variables used in the Study

##### 1. Dependent Variables

Overall reading and science proficiency scores represented the students' level of mastery of certain skills on the cognitive tests of reading ability and science ability, respectively. The proficiency scores provide a means of distinguishing total scores and score gains, as measured by overall IRT-Estimated Number Right scores and the norm-referenced t-scores, from criterion-referenced measurements of specific skills. At several points along the score scale of the reading test, four-item clusters of test questions having similar content and difficulty were identified. A student was assumed to have mastered a particular level of proficiency if at least three of the four items in the cluster were answered correctly, and to have failed at this level if two or more items were wrong. Clusters of items provide a more reliable test of proficiency than do single items because of the possibility of guessing in a multiple-choice test. Additionally, the proficiency levels followed a Gutman model, that is, a student passing at a particular skill level was expected to have mastered all lower levels. Conversely, failure at a given level indicated non-mastery at higher levels (U.S. Department of Education, 1994). Scoring for both reading and science proficiency was 1 = below level 1, 2 = level 1, but not level 2, and 3 = level 2. Higher scores indicate greater reading and science proficiency. Overall math



proficiency was similarly partitioned. However, this variable was partitioned into 4 exhaustive and mutually exclusive categories, below level 1, level 1, level 2, and level 3.

Three dichotomously coded response variables that represent whether or not the student exhibited math proficiency, reading proficiency, and science proficiency, respectively at the highest level were included in the discriminant analyses. In addition, the standardized scores on the math, reading, and science cognitive tests were included for the final phase of the analyses.

## 2. Independent Variables

The educational productivity factors were derived from Walberg's Model of Educational Productivity (Walberg, 1984). Self-concept was measured using the composite of items in multi-part question from the student questionnaire. These items were "I feel good about myself", "I feel I am a person of worth, the equal of other people", "I am able to do things as well as most other people", "On the whole, I am satisfied with myself", "I certainly feel useless at times", "At times I think I am no good at all", and "I feel I do not have much to be proud of". Each of the items was standardized separately to a mean of 0 and standard deviation of 1. The composite represents the sum of the standardized scores. Some of the items had to be reverse coded so that lower scores corresponded to lower self-concept and higher scores corresponded to higher self-concept.

Locus of control was also measured using a composite of items. These items were "I don't have enough control over the direction my life is taking", "In my life, good luck is more important than hard work for success", "Every time I try to get ahead, something or somebody stops me", "My plans hardly ever work out, so planning only

makes me unhappy”, “When I make plans, I am almost certain I can make them work”, and “Chance and luck are very important for what happens in my life”. The items were standardized and recoded so that lower scores corresponded to lower control.

Socio-economic status (SES) was constructed using the following items: father’s education level, mother’s education level, father’s occupation, mother’s occupation, and family income. The items were summed and then standardized to form the composite measure of SES. The other predictors were either derived from transcript data or the self-administered questionnaire.

Other independent variables included in the present study were parent’s education, post-secondary education plans, advanced, enriched, or accelerated course work in the subject areas English, science, and math, respectively, English, math, and science grades between 6<sup>th</sup> and 8<sup>th</sup> grades, and number of hours spent on homework each week.

## **Results**

The results of the statistical analyses are included in the Appendix.

### **I. Canonical Correlation Procedure**

#### **A. Within Group Correlation Structure**

##### **1) Academic Proficiency Variables**

The correlation between math and reading proficiency was  $r = .46$ . The correlation between math and science proficiency was  $r = .54$ , and the correlation between reading and science was  $r = .42$ . Each of these correlations was found to be moderately or highly significant.

##### **2) Productivity Factors**

I will comment on some of the moderately to highly correlated factors (i.e.,  $r \geq .4$ ). Advanced math was correlated with Advanced English ( $r = .52$ ) and advanced science ( $r = .42$ ). Advanced English was correlated with advanced science ( $r = .60$ ). Post-secondary education plans was moderately related to parent's education ( $r = .40$ ) and socio-economic status ( $r = .42$ ). Locus of control was related to self-concept ( $r = .54$ ). Parent's education and socio-economic status were highly correlated ( $r = .86$ ). Finally, science grades were related to English grades ( $r = .47$ ).

#### B. Between Groups

None of the productivity factors appeared highly related to the academic proficiency response variables. However, this is probably due to the high intercorrelations among the productivity factors. Yet some of the productivity factors were more related to the proficiency variables than others. Advanced math, math grades, post-secondary education plans, socio-economic status, parent's education, English grades, and science grades appear to be moderately related to both math and science proficiency scores. On the other hand, only parent's education, socio-economic status, English, and science grades were at least moderately related to reading proficiency. It is somewhat surprising that advanced English was not, at least moderately related to the reading proficiency.

#### C. Squared Canonical Correlations

Three canonical correlations were produced in the current analysis. The values of these correlations were .61, .30, and .19. The squares of these correlations were .37, .09, and .04. The squares of the correlations represent the proportion of variance of the academic proficiency variables that is explained by the productivity factors. Thus, the

first square correlation coefficient,  $\lambda_1 = .37$ , indicates that 37% of the variance of the Proficiency variables can be attributed to the productivity factors. The square root of the squared canonical correlation represents the correlation between pairs of canonical variables. There are 6 possible canonical variables (3 for the response variables and 3 for the predictor variables), thus there are 3 canonical correlations.

Significance Testing. Wilk's Lambda,  $\Lambda = .55$ , for the test of whether the *first, second, and third canonical correlations were all zero* was significant,  $F = 5.38, p < .05$ . This result suggests that at least one of the canonical correlations was non-zero. The second Wilk's Lambda,  $\Lambda = .87$ , for the test of whether the *second and third canonical correlations were both zero* was also significant,  $F = 1.83, p < .05$ . This result indicates that at least the second canonical correlation is non-zero. Based on the third value for Wilk's Lambda, the third canonical correlation was *no different from zero*,  $F = 1.11, ns$ .

#### D. Equations of the Canonical Variables

Using the standardized canonical coefficients, the equations of the six canonical variables were:

$$\begin{aligned}
y_1^* &= .60y_1 + .22y_2 + .39y_3 \\
y_2^* &= -.64y_1 + 1.12y_2 + -.20y_3 \\
y_3^* &= -.89y_1 + -.16y_2 + 1.14y_3 \\
x_1^* &= .28x_1 + .10x_2 + -.23x_3 + .20x_4 + .19x_5 + .03x_6 + .24x_7 + -.03x_8 + .10x_9 + .21x_{10} + .23x_{11} + .28x_{12} \\
x_2^* &= -.32x_1 + .12x_2 + .10x_3 + -.51x_4 + -.38x_5 + .54x_6 + .27x_7 + .39x_8 + .05x_9 + -.23x_{10} + .74x_{11} + -.12x_{12} \\
x_3^* &= .36x_1 + -.69x_2 + .39x_3 + -.57x_4 + -.23x_5 + .21x_6 + .48x_7 + .42x_8 + -.09x_9 + .19x_{10} + -.17x_{11} + .07x_{12} ,
\end{aligned}$$

where  $y_1$  = Math Proficiency,  $y_2$  = Reading Proficiency,  $y_3$  = Science Proficiency,  
 $x_1$  = advanced math,  $x_2$  = advanced English,  $x_3$  = advanced science,  $x_4$  = math grades,  
 $x_5$  = post - secondary education plans,  $x_6$  = locus of control,  $x_7$  = parent's education,  
 $x_8$  = self - concept,  $x_9$  = socio - economic status,  $x_{10}$  = homework,  
 $x_{11}$  = English grades,  $x_{12}$  = science grades.

The greatest weight for the first canonical variable of the proficiency variables ( $y_1^*$ ) is attributed to math proficiency (.60) and the greatest weights for the corresponding canonical variable of the productivity factors ( $x_1^*$ ) is attributed to both advanced math (.28) and science grades (.28). The most weight for the second pair of equations ( $y_2^*$  and  $x_2^*$ ) is attributed to reading proficiency (1.12) and English grades (.74), and the most weight for the third pair of equations ( $y_3^*$  and  $x_3^*$ ) can be attributed to science proficiency (1.14) and parent's education (.48). Thus, the first pair of canonical variables refers mainly to the relationship between math proficiency and advanced math and science grades. The second pair of canonical variables refers mainly to the relationship between reading proficiency and English grades. The third pair refers primarily to the relationship between science proficiency and parent's education.

The canonical structure reveals that math and science proficiency are highly correlated with their first canonical variable ( $r = .91$  and  $.80$ , respectively). Reading is at least moderately correlated with this variable ( $r = .66$ ). The second canonical variable

appears to be highly associated with reading proficiency ( $r = .74$ ). And the third canonical variable seems to be most related to science proficiency ( $r = .59$ ).

The first canonical variable for the educational productivity factors is related to science grades ( $r = .69$ ) most predominantly, but also English grades ( $r = .62$ ), socioeconomic status ( $r = .66$ ), parent's education ( $r = .66$ ), post-secondary education plans ( $r = .60$ ), math grades ( $r = .56$ ), and advanced math ( $r = .47$ ). The second variable is most related to English grades ( $r = .53$ ). And the third canonical variable is most related to self-concept ( $r = .49$ ).

#### E. Redundancy Analysis

Analyses of the squared correlations between the academic proficiency variables and the canonical variables of the productivity factors indicates that the first canonical variable accounts for 31% ( $.5549^2$ ) of the variance of math proficiency, 16% of the variance of reading proficiency ( $.4048^2$ ), and 24% of the variance of science proficiency ( $.4904^2$ ). The second canonical variable accounts for less than 1% of the variance for both math ( $-.0704^2$ ) and science ( $-.0223^2$ ), but about 5% for reading ( $.2251^2$ ). The third canonical variable accounts for less than 1% of the variance for math and reading, and only 1.3% for science. Thus, the first canonical variable accounts for the majority of the variance of the three proficiency measures.

On the basis of the correlations between the proficiency measures and the productivity factors with their respective canonical variables as well as the relationship between the canonical variables themselves, it appears that the proficiency variables are related to several factors. These factors include primarily advanced math, math, English,

and science grades, post-secondary education plans, parent's education, and socioeconomic status.

## II. Discriminant Analysis

Three stepwise discriminant analyses were conducted to determine a subset of the productivity factors that best distinguished the categories of the response variables (e.g., proficiency or non-proficiency at the highest possible level). For each subject area, proficiency levels were dichotomously scored with 1 representing proficiency and 0 representing non-proficiency.

### 1. Math Proficiency

#### A. Univariate Tests

The univariate tests indicated that three of the twelve-discriminant variables showed no difference in average scores between the proficient and non-proficient groupings. These variables included advanced science ( $F = .61, p = .43$ ), locus of control ( $F = 3.72, p = .05$ ), and self-concept ( $F = 1.99, p = .16$ ). The mean scores of the remaining nine factors were different between the two levels of math proficiency.

#### B. Multivariate Test

The Mahalanobis distance between the centroids for the math proficient and non-proficient groups was significant,  $D^2 = 1.90, F = 9.90, p < .0001$ . Thus, in general there was a difference in the set of discriminant variables between the proficient and non-proficient groups.

#### C. Discriminant Function

The discriminant function was calculated by normalizing the raw coefficients of the discriminant variables. The discriminant function for math proficiency is

$$t = .802x_1 + .093x_2 + -.445x_3 + .199x_4 + .085x_5 + -.153x_6 + .089x_7 + \\ -.029x_8 + .0001x_9 + .166x_{10} + .045x_{11} + .203x_{12} ,$$

where  $x_i = X_i - \bar{X}_i$  for  $i = 1, \dots, 12$ .

Here  $x_1$  = Advanced Math,

$x_2$  = Advanced English,

$x_3$  = Advanced Science,

$x_4$  = Math Grades,

$x_5$  = Post -secondary Education Plans,

$x_6$  = Locus of Control,

$x_7$  = Parent's Education,

$x_8$  = Self - Concept,

$x_9$  = Socio - economic Status,

$x_{10}$  = Homework,

$x_{11}$  = English Grades,

$x_{12}$  = Science Grades.

The standardized pooled within-class canonical coefficients indicate that advanced math (.58) had the greatest influence on the discriminant function, followed by homework hours (.35), science grades (.31), advanced science (-.30), math grades (.29), parent's education (.17), socio-economic status (.17), post-secondary education plans (.15), locus of control (-.14), advanced English (.06), English grades (.06), and self-concept (-.03).

#### D. Misclassification Analysis

The classification summary showed that 25% of the proficient observations were misclassified and 25% of the non-proficient observations were misclassified by the discriminant function. Overall then, 25% of the observations were misclassified.

#### E. Stepwise Procedure

According to the results of the stepwise discriminant analysis procedure, the final subset of variables included in the discriminant function included advanced math, socio-



economic status, math grades, homework, science grades, and advanced science. The partial  $R^2$  statistics for the variables indicated that advanced math explained 9.7% of the variance of math proficiency, socio-economic status explained 7.8%, math grades explained 3.7%, homework explained 2.8%, science grades explained 1.8% and advanced science explained 1.3% of the variance. Each of these proportions was significant at  $\alpha = .05$ . The Wilks' Lambda for the final model was .755,  $F = 4.529$ ,  $p < .0001$ . The average squared canonical correlation was .25, indicating that the discriminant function including the given subset of factors explained 25% of the variance of math proficiency.

## 2. Reading Proficiency

### A. Univariate Tests

The univariate tests indicated that there were differences in the mean scores for advanced math ( $F = 5.56$ ,  $p = .02$ ), advanced English ( $F = 4.03$ ,  $p = .05$ ), math grades ( $F = 11.16$ ,  $p = .0009$ ), education plans ( $F = 12.14$ ,  $p = .0006$ ), locus of control ( $F = 4.96$ ,  $p = .03$ ), parent's education ( $F = 31.32$ ,  $p = .0001$ ), SES ( $22.57$ ,  $p = .0001$ ), homework ( $F = 5.07$ ,  $p = .02$ ), English grades ( $F = 32.21$ ,  $p = .0001$ ), and science grades ( $F = 30.91$ ,  $p = .0001$ ). No difference was found for the factors advanced science and self-concept.

### B. Multivariate Test

The Mahalanobis distance between the centroids for the reading proficient and non-proficient groups was  $D^2 = .91$ ,  $F = 6.06$ ,  $p = .0001$ . Therefore, there was an overall difference in the means of the discriminant variables between the proficient and non-proficient groups.

### C. Discriminant Function

$$t = .061x_1 + .492x_2 + -.242x_3 + .075x_4 + -.008x_5 + .257x_6 + .490x_7 + \\ -.535x_8 + -.206x_9 + .128x_{10} + .455x_{11} + .400x_{12},$$

where  $x_i = X_i - \bar{X}_i$  for  $i = 1, \dots, 12$ .

The standardized pooled within-class canonical coefficients indicate that parent's education had the greatest influence on the discriminant function (.58), followed by English grades (.40), science grades (.39), self-concept (-.34), advanced English (.22), homework (.18), locus of control (.15), SES (-.12), advanced science (-.10), math grades (.07), advanced math (.03), and education plans (-.01).

#### D. Misclassification Analysis

The classification summary showed that 28% of the non-proficient subjects were misclassified and 29% of the proficient observations were incorrectly classified. The overall proportion of misclassified subjects was 29%.

#### E. Stepwise Procedure

The results of the stepwise discriminant procedure indicated that English grades, parent's education, science grades, self-concept, and homework were included in the final subset for the discriminant function. The partial  $R^2$  statistics for the variables indicated that English grades explained 8% of the variance of reading proficiency, parent's education explained 6%, science grades explained 2%, and self-concept and homework both explained 1%. The contribution of homework hours was not significant at  $\alpha = .05$ . Wilks' Lambda was .83,  $F = 14.035$ ,  $p < .0001$ . The average squared canonical correlation was .17, indicating that 17% of the variance of reading proficiency was explained by the subset of variables included in the discriminant function.

### 3. Science Proficiency

### A. Univariate Tests

The univariate tests indicated that there was a difference in mean scores for advanced math ( $F = 11.63, p = .0007$ ), math grades ( $F = 11.72, p = .0007$ ), post-secondary education plans ( $F = 25.23, p = .0001$ ), locus of control ( $F = 4.04, p = .05$ ), parent's education ( $F = 29.06, p = .0001$ ), SES ( $F = 24.70, p = .0001$ ), homework ( $F = 7.27, p = .007$ ), English grades ( $F = 16.46, p = .0001$ ), and science grades ( $F = 19.06, p = .0001$ ).

### B. Multivariate Test

The Mahalanobis distance between the centroids for the science proficient and non-proficient groups was significant,  $D^2 = .95, F = 4.91, p < .0001$ . Thus, there are differences in the discriminant variables across the levels of science proficiency.

### C. Discriminant Function

The discriminant function was determined by normalizing the raw coefficients of the discriminant variables. The discriminant function for science proficiency was

$$t = .713x_1 + -.214x_2 + -.300x_3 + .159x_4 + .255x_5 + -.200x_6 + .336x_7 + .008x_8 + -.0000001x_9 + .151x_{10} + .229x_{11} + .171x_{12},$$

where  $x_i = X_i - \bar{X}_i$  for  $i = 1, \dots, 12$ .

The standardized pooled within-class canonical coefficients reveal that parent's education (.44) had the most influence on the discriminant function, followed by advanced math (.37), education plans (.32), homework (.23), English grades (.23), science grades (.19), math grades (.16), advanced science (-.14), locus of control (-.13), advanced English (-.10), self-concept (.06), and SES (-.01).

### D. Misclassification Analysis

Thirty-four percent of the non-proficient observations were misclassified as proficient and 31% of the proficient observations were misclassified as non-proficient. Thirty-two percent overall were misclassified by the discriminant function.

#### E. Stepwise Procedure

The stepwise procedure was used to determine the best model to distinguish proficient from non-proficient observations. The results of this procedure indicated that the best model included the factors parent's education, post-secondary education plans, advanced math, English grades, and homework. The partial correlation coefficients for these variables indicated that parent's education explained 8% of the variance of science proficiency, post-secondary plans explained 3% of the variance, advanced math explained 2%, and English grades and homework both explained about 1%. Wilks' Lambda statistic was .86,  $F = 10.888$ ,  $p < .0001$ . The average squared canonical correlation was .14, indicating that 14% of the variance of science proficiency was explained by the subset of variables included in the discriminant function.

### III. Principal Component and Factor Analyses

Before considering the results of the principal component analysis and factor analysis, it is important to examine the interrelationships among the variables as depicted in the correlation matrix.

The correlation structure reveals that the variables representing advanced course work in English, math, and science were related with correlations ranging from  $r = .49$  to  $.61$ . Grades in the three subject areas were also correlated with one another more so than with the other measures. The correlations among the "grade" variables ranged from  $r = .38$  to  $.46$ . Locus of control and self-concept appear to be correlated with one another,  $r =$

.54. Socio-economic status, parent's education, and post-secondary plans are correlated, with correlations ranging from  $r = .39$  to  $.85$ . The variable homework does not appear to be strongly correlated with any particular variable. Thus, the initial correlation structure among the variables suggests that there may be three to four underlying constructs within the set of independent variables.

#### A. Principal Component Analysis

The principal component analysis output indicates that 4 components have eigenvalues greater than 1. The eigenvalues represent the variances of the component scores. The first component accounts for 28% of the variance of the original 12 variables. The second component accounts for 16%, the third accounts for 12% and the fourth component accounts for 9% of the variance of the original twelve variables. Hence, these four components together explain 65% of the variance of the original variables.

As indicated from the factor loadings, English grades (.34), math grades (.30), science grades (.36), SES (.34), parent's education (.33), and post-secondary education plans (.36) have relatively high loadings on the first component. Advanced English (.54), advanced science (.54), and advanced math (.45) have high loadings on the second component. Locus of control (.46, .37) and self-concept (.48, .49) had high loadings on both third and fourth components. However, as suggested by the correlation structure, the three variables representing grades and the three variables representing family/background do not appear to represent the same dimension, even though all six of these loaded onto the first component.

The results of this exploratory procedure suggest that the twelve-variable model may be reduced to a lesser dimension (at least three). Factor analysis was applied to the data to determine more accurately whether or not the variables represent separate dimensions.

#### B. Factor Analysis

Again, before considering the results of the factor analysis explicitly, it is noteworthy to observe the partial correlations between the independent variables. The partial correlations among the variables, particularly, the variables representing “grades” and those representing family/background, shows that math, English, and science grades have relatively high correlations among one another, and the family/background variables have similarly high correlations among one another. The partial correlations between these subgroups were much smaller.

Using exploratory techniques, with no preset number of factors, the procedure retained 3 factors based on the proportion criterion. The eigenvalues for the first three factors were 2.77, 1.46, and .89. The first factor extracts 61% of the variance from the total common variance of the variables. The second factor extracts 32% and the third factor extracts 19% of the common variance. Many of the eigenvalues are negative. This explains why the sum of the variance extracted by the first three factors exceeds 100% of the common variance.

The factor pattern suggests that SES, parent’s education, post-secondary education plans, science grades, English grades, math grades, and locus of control have high loadings on the first factor. Advanced English, science, and math have high

loadings on the second factor. Finally, locus of control and self-concept have high loadings on the third variable.

Varimax rotation was applied to further reduce the ambiguity associated with the factor loadings. The loadings on the factors after the varimax rotation were as follows: locus of control, science grades, self-concept, English grades, math grades, and post-secondary education plans had relatively high loadings on factor 1. SES and parent's education had high loadings on factor 2. Finally, advanced course work in English, science, and math had high loadings on the third factor. Homework hours did not have a notably high loading on any of these three factors. Not surprisingly, homework had a very small final communality estimate, .07. This estimate indicates that 7% of the variance of the variable homework was extracted by the common factors. Thus, 93% of the variance of this variable must be attributed to the unique factor corresponding to the homework variable.

From these analyses, it is clear that certain subsets of the original 12 variables may represent separate dimensions as predictors of academic achievement. Based on the results of the factor analyses and the correlations among the variables, 4 "new" composite variables were created. The new variables were: "motivation", "grades", "advanced course work", and "family background". Motivation was created by adding the scores of the variables locus of control and self-concept, grades represented the sum of the scores from English, math, and science grades, advanced course work was the sum of the scores of the variables representing advanced course work in math, English, and science, and family background was the sum of the scores on the variables SES, post-secondary education plans, and parent's education. Homework was not included in any new scale.

The new variables, as well as homework, gender and race are used in the subsequent MANOVA procedure.

#### IV. MANOVA

Before conducting multivariate analysis of variance (MANOVA), the new variables were scaled into mutually exclusive and exhaustive levels. In addition, a rescaled version of the variable representing the number of hours of homework done per week, "homework", was created having only 3 levels. The new scale for the variable homework was as follows: 1 = 0 to 1.99 hours, 2 = 2 to 10.49 hours, and 3 = 10.50 hours and up. Motivation and family background were set up with only two levels: 0 = low and 1 = high. These levels were determined by finding the mean of the variables, then reassigning the observations to "low" if the score was below the mean and "high" if the score was above the mean.

For the variables grades and advanced course work, a somewhat different procedure was applied to assign levels for these variables. Recall, a new variable was created to represent the sum of the variables English, math, and science grades. The coding scheme for the original subject grade variables was as follows: 5 = Mostly A's, 4 = Mostly B's, 3 = Mostly C's, 2 = Mostly D's, and 1 = Mostly below D. So, a score of 15 on the new variable (sum of the three subject-grade scales) represented mostly A's for all three-subject areas and a score of 3 represented mostly below D for all three subject areas. Thus, the levels of the variable grades were determined as follows: 0 = factor score of 3; 1 = factor score less than or equal to 6 and greater than 3; 2 = factor score less than or equal to 9 and greater than 6; 3 = factor score less than or equal to 12 and greater



than 9; and 4 = factor score greater than 12, where factor score represents the values of the factor representing the sum of the three original subject-grade variables.

For advanced course work a similar procedure was used to determine its levels. For the three original advanced course work variables, 1 = yes, advanced course work and 0 = no advanced course work. Thus, the values for the new variable representing the sum of advanced course work in the three subject areas ranged from 0 = no advanced course work to 3 = advanced course work in all three subjects. The levels for the variable advanced course work were assigned as follows: 0 = factor score equal 0; 1 = factor score equal 1; 2 = factor score equal 2; and 3 = factor score equal 3, where factor score represents the values of the factor representing the sum of the three original course work variables.

The independent variables included grades, motivation, advanced course work, family background, homework, race, and gender. The three dependent variables were the raw scores on the cognitive tests in the three subjects, reading, math, and science.

#### A. Analysis of Variance Procedures

The first part of the MANOVA examined whether the means of the dependent variables differed according to the levels of the independent variables. The results of this part of the analysis indicated that math standardized scores differed according to grades ( $F = 16.17, p = .0001$ ), family background ( $F = 17.63, p = .0001$ ), homework ( $F = 4.46, p = .0122$ ), and race ( $F = 5.42, p = .0003$ ). Reading scores differed according to grades ( $F = 8.98, p = .0001$ ), family background ( $F = 10.75, p = .0011$ ), and gender ( $F = 4.71, p = .0307$ ). Science scores differed according to grades ( $F = 12.79, p = .0001$ ), family background ( $F = 15.08, p = .0001$ ), homework ( $F = 7.13, p = .0009$ ), race ( $F = 3.11, p =$

.0155), and gender ( $F = 5.42, p = .0204$ ). Scores for math, reading, and science were no different across levels of advanced course work or motivation. Additionally, tests of interactions between race and the other significant factors indicated that the differences in the mean-scores of math, reading, and science outcomes with respect to groupings based on grades, family background, and homework remained stable across different racial groups.

#### B. Multivariate Analysis of Variance

The second part of the analysis examined if there were differences between the centroids of the groups or levels of independent variables. The centroids are the vectors of the means of the scores on the cognitive tests for reading, math, and science, respectively for each level of the independent variables. There were differences in the centroids at different levels of the variables grades ( $\Lambda = .83, F = 5.80, p = .0001$ ), family background ( $\Lambda = .95, F = 6.66, p = .0002$ ), homework ( $\Lambda = .95, F = 3.16, p = .0046$ ), race ( $\Lambda = .92, F = 2.41, p = .0044$ ), and gender ( $\Lambda = .91, F = 11.74, p = .0001$ ). On the other hand there were no differences in the centroids for the variables motivation ( $\Lambda = .99, F = 1.03, p = .3785$ ) and advanced course work ( $\Lambda = .97, F = 1.06, p = .3881$ ). These results suggest that at least one pair of centroids among the levels of the variables grades, family background, homework, race, and gender differed. The Wilks' Lambda statistic  $\Lambda$ , is defined as the ratio of the determinant for the within-group variations  $W$ , to the determinant of the total variation and covariation matrix  $T$  ( $T = W + B$ ). Specifically,

$$\Lambda = \frac{|W|}{|W + B|},$$

where  $W$  = matrix of with-group variations and  $B$  = matrix of between-group variations and covariations. This ratio then indicates the proportion of variance that

can be accounted for by the within-group variations. Thus, larger values for the Wilks' statistic reveal that there is little difference in the effect of the different levels of the independent variables on the dependent variable.

### C. Pair-wise Comparisons

Based on the results of the Analysis of Variance and Multiple Analysis of Variance procedures, pair-wise comparisons were conducted using Scheffe's multiple comparison option for variables having a significant F-value. According to Kirk (1982), Scheffe's procedure can be used to evaluate a posteriori contrasts among means when the cell sizes are unequal. Additionally, this test is robust with respect to non-normality and heterogeneity of variance. The disadvantage is that the Scheffe test is less powerful than other tests for pair-wise comparisons (Kirk, 1982).

#### 1) Grades

For the variable grades, the results indicate that the mean math scores for group 4 differed from groups 3, 1, and 2 and group 3 differed from group 2. The average reading scores of group 4 differed from groups 3 and 2 and group 3 differed from group 2. The mean for science scores differed between group 4 and groups 3 and 2. These results suggest that the centroid difference between the groups is primarily due to the difference between group 4 and the other groups. In particular, the results indicate that students receiving higher grades in math, English, and science perform better on cognitive tests in these subject areas.

#### 2) Family

There were differences between the high and low levels of family background for all three dependent measures, with the mean scores for those in the high group out-performing students in the low group.

### 3) Homework

A comparison of the levels of the amount of homework done during the week indicates that math scores differ between those who do at least 10.50 hours and those who do 2 to 10.49 hours of homework per week as well as those who do 0 to 1.99 hours. Student's who do more than 10.50 hours have high mean reading scores than those who do between 2 and 10.49 hours. The mean science score for those who do at least 10.50 hours of homework is greater than those who do between 2 and 10.49 hours as well as those who do less than 1.99 hours of homework per week. These results suggest that students who do more homework tended to do better on the subject tests than those who did less homework.

### 4) Race

Comparisons of the average cognitive test scores by race indicate that Asian-Pacific Islanders (API) out-performed Native Americans, African Americans, and Hispanics in math, and API's out-performed Hispanics and African Americans in reading and science. Whites had higher average scores on all three tests than both African Americans and Hispanics. On the other hand, Native Americans, African American, and Hispanics all did relatively equal on the three subject tests.

### 5) Gender

Males out-performed females in science, but females out-performed males in reading. There was no difference between males and females in math.

## **Discussion**

The current investigation examined the relationship between several educational productivity variables and achievement outcomes in three separate subject areas, math, reading, and science. Specifically, the study made use of various multivariate statistical technique to determine the nature and magnitude of the relationship between advanced and enriched course work, grades, future education plans, parental education, socio-economic status, self-concept, locus of control, and homework time and achievement in reading, math, and science.

The results indicated that many of the variables were intercorrelated with one another but some subsets of the predictor variables exerted more influence on the achievement outcomes than others. Overall, the productivity factors explained over 50% of the variance of the achievement outcomes. In particular, the results of the canonical correlation analysis suggest that the achievement outcomes in the three subject areas are most related to advanced course work in English, math, and science, grades in science classes, post-secondary education plans, parent's education, and socio-economic status. Yet, it appears that the relationship between math achievement and these predictor variables had the greatest influence on the overall association between the two sets of variables.

Three separate discriminant analysis were conducted to determine which subsets of the Productivity factors were most influential in distinguishing between proficiency and non-proficiency in the three subject areas. For math proficiency, advanced math course work discriminated best between proficient and non-proficient. Results of the stepwise discriminant procedure agreed with the previous outcome. Because math is a

cumulative subject, it is not surprising that exposure to higher level or advanced course work in math has the most impact on determining whether or not a student will exhibit proficiency. Analysis of the coefficients of the discriminant function for reading proficiency suggests that parent's education followed by grades in English classes discriminated best between proficiency and non-proficiency. The stepwise procedure supported this finding, indicating that English grades, followed by parent's education had the greatest impact on the reading proficiency variable. These results reveal that more educated parent's are more inclined to encourage their children to read and thus these student may show greater proficiency in reading. Students' grades in English also seem to reflect their potential for proficiency in reading. For science proficiency, parent's education had the greatest impact on distinguishing between proficiency and non-proficiency. This result shows that more educated parents tend to encourage their children to perform better in science. This is probably manifested through parents' encouragement to read and devote more time to science projects and class work.

Principal component and factor analyses were applied to the data to explore the interrelationships among the productivity variables. During the course of the procedures, it was revealed that combining some of the variables could reduce the dimension of the Productivity model. In particular, four new variables were created based on the loadings of the variables on certain common factors. These included motivation, which represented the sum of the variables locus of control and self-concept; grades, representing the sum of the grades in English, math, and science; advanced course work, comprised of the sum of the responses to advanced course work in the three subject areas; and family background, which was the sum of parent's education, post-secondary

education plans, and SES. The “new” variables were then rescaled into distinct levels for use in the subsequent multivariate analysis of variance procedure.

The MANOVA procedure revealed that average math achievement differed according to grades, family background, homework, and race. Reading achievement differed according to grades, family background, and gender. Science achievement varied according to grades, family background, homework, race, and gender. Multivariate tests suggested that the average achievement in all three subject areas under study differed for different levels of grades, family background, homework, race, and gender. Both of these analyses indicate that differences in achievement in reading, math, and science are primarily due to differences in levels of course grades, family background variables, and time spent on homework. It should also be pointed out that the effect of the individual variables, particularly advanced course work in the different subject areas, on achievement outcomes may have been compromised by combining these variables into a single factor.

Pair-wise analyses suggested that student’s receiving higher grades tended to do better in the three subject areas than those who did not. Student’s having more educated parents, higher aspirations for future education, and from more socio-economically advantaged backgrounds tended to also do better in the three subject areas. Student’s who spent the maximum number of hours on homework out-performed those who spent much less time on homework across all three subject areas. As for race, API’s and Whites out-performed African Americans, Hispanics, and Native Americans in reading, math, and science. There were however, no significant differences in performance between API’s and Whites, or between African Americans, Hispanics, and Native

Americans. Finally, males tended to out-perform females in science, but females did better than males in reading. There was no gender difference in math achievement.

In conclusion, the current study indicates that several factors are related to achievement outcomes. Furthermore, certain variables more related to particular subject areas than others. Math achievement seems to be effected by advanced math course work than the other variables. Reading appears to be more effected by previous grades in English and parent's education. Finally, science achievement seems to be influenced most by parent's education level. On the other hand, certain common factors seem to have a significant influence on academic achievement. These factors include grades, family background, and homework, as well as race and gender. Although gender, race, and education of parent's represent preexisting conditions, grades and homework are alterable conditions, that if maximized appear to increase the likelihood of higher performance in all three subject areas.



## References

- Cody, R.P., & Smith, J.K. (1997). Applied Statistics and the SAS Programming Language (4<sup>th</sup> ed). Upper Saddle River, NJ: Prentice Hall.
- Kirk, R.E. (1982). Experimental Design: Procedures for the Behavioral Sciences (2<sup>nd</sup> ed). Pacific Grove, CA: Brooks/Cole.
- Peng, S., & others. (1995). Understanding racial-ethnic differences in secondary school science and mathematics achievement: Research and development report (Doc. No. SE 055 976). Washington, DC: National Center for Education Statistics. (ERIC Document Reproduction Service No. ED 381 342).
- Tacq, J. (1997). Multivariate Analysis Techniques in Social Science Research: From Problem to Analysis. Thousand Oaks, CA: Sage.
- U.S. Department of Education. (1994). National Educational Longitudinal Study of 1988. Second Follow-Up: Student Component Data File User's Manual. Washington, DC: National Center for Education Statistics, U.S. Department of Education.
- U.S. Department of Education. (1989). National Educational Longitudinal Study of 1988. Base Year: Student Component Data File User's Manual. Washington, DC: National Center for Education Statistics, U.S. Department of Education.
- Walberg, H.J. (1984). Improving the productivity of America's schools. Educational Leadership, May, 19-27.

BEST COPY AVAILABLE



U.S. Department of Education  
Office of Educational Research and Improvement (OERI)  
National Library of Education (NLE)  
Educational Resources Information Center (ERIC)



TM034700

# REPRODUCTION RELEASE

(Specific Document)

## I. DOCUMENT IDENTIFICATION:

Title: *An Examination of the Relationship Between Learning Variables and Academic Achievement Outcomes*

Author(s): *John P. Thomas, Ph.D.*

Corporate Source:

*College of Lake County Illinois*

Publication Date:

## II. REPRODUCTION RELEASE:

In order to disseminate as widely as possible timely and significant materials of interest to the educational community, documents announced in the monthly abstract journal of the ERIC system, *Resources in Education* (RIE), are usually made available to users in microfiche, reproduced paper copy, and electronic media, and sold through the ERIC Document Reproduction Service (EDRS). Credit is given to the source of each document, and, if reproduction release is granted, one of the following notices is affixed to the document.

If permission is granted to reproduce and disseminate the identified document, please CHECK ONE of the following three options and sign at the bottom of the page.

The sample sticker shown below will be affixed to all Level 1 documents

The sample sticker shown below will be affixed to all Level 2A documents

The sample sticker shown below will be affixed to all Level 2B documents

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

*Sample*

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

1

Level 1



Check here for Level 1 release, permitting reproduction and dissemination in microfiche or other ERIC archival media (e.g., electronic) and paper copy.

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE, AND IN ELECTRONIC MEDIA FOR ERIC COLLECTION SUBSCRIBERS ONLY, HAS BEEN GRANTED BY

*Sample*

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

2A

Level 2A



Check here for Level 2A release, permitting reproduction and dissemination in microfiche and in electronic media for ERIC archival collection subscribers only

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY

*Sample*

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

2B

Level 2B



Check here for Level 2B release, permitting reproduction and dissemination in microfiche only

Documents will be processed as indicated provided reproduction quality permits.  
If permission to reproduce is granted, but no box is checked, documents will be processed at Level 1.

I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce and disseminate this document as indicated above. Reproduction from the ERIC microfiche or electronic media by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction by libraries and other service agencies to satisfy information needs of educators in response to discrete inquiries.

Signature: *John P. Thomas* Printed Name/Position/Title: *John P. Thomas / Prof. Mathematics*  
Organization/Address: *19351 W. Washington; Grayslake, IL* Telephone: *(847) 543-2912* FAX:   
E-Mail Address: *eng512@clc.cc.il.us* Date: *12-18-02*

Sign here, please



(over)

### III. DOCUMENT AVAILABILITY INFORMATION (FROM NON-ERIC SOURCE):

If permission to reproduce is not granted to ERIC, or, if you wish ERIC to cite the availability of the document from another source, please provide the following information regarding the availability of the document. (ERIC will not announce a document unless it is publicly available, and a dependable source can be specified. Contributors should also be aware that ERIC selection criteria are significantly more stringent for documents that cannot be made available through EDRS.)

Publisher/Distributor:
Address:
Price:

### IV. REFERRAL OF ERIC TO COPYRIGHT/REPRODUCTION RIGHTS HOLDER:

If the right to grant this reproduction release is held by someone other than the addressee, please provide the appropriate name and address:

Name:
Address:

### V. WHERE TO SEND THIS FORM:

<p>Send this form to the following ERIC Clearinghouse:</p> <p><b>ERIC CLEARINGHOUSE ON ASSESSMENT AND EVALUATION</b> <b>UNIVERSITY OF MARYLAND</b> <b>1129 SHRIVER LAB</b> <b>COLLEGE PARK, MD 20742-5701</b> <b>ATTN: ACQUISITIONS</b></p>
---

However, if solicited by the ERIC Facility, or if making an unsolicited contribution to ERIC, return this form (and the document being contributed) to:

**ERIC Processing and Reference Facility**  
4483-A Forbes Boulevard  
Lanham, Maryland 20706

Telephone: 301-552-4200  
Toll Free: 800-799-3742  
FAX: 301-552-4700  
e-mail: [info@ericfac.piccard.csc.com](mailto:info@ericfac.piccard.csc.com)  
WWW: <http://ericfacility.org>